The dynamics of Pythagorean triples

Dan Romik *

Abstract

We construct a piecewise onto 3-to-1 dynamical system on the positive quadrant of the unit circle, such that for rational points (which correspond to normalized Primitive Pythagorean Triples), the associated ternary expansion is finite, and is equal to the address of the PPT on Barning's [9] ternary tree of PPTs, while irrational points have infinite expansions. The dynamical system is conjugate to a modified Euclidean algorithm. The invariant measure is identified, and the system is shown to be conservative and ergodic. We also show, based on a result of Aaronson and Denker [2], that the dynamical system can be obtained as a factor map of a cross-section of the geodesic flow on a quotient space of the hyperbolic plane by the group $\Gamma(2)$, a free subgroup of the modular group with two generators.

AMS 2000 Mathematics Subject Classification: 37A45

1 Introduction

The starting point of this paper is a theorem, attributed to Barning [9], on the structure of the set of Primitive Pythagorean Triples, or PPTs. Recall that a PPT is a triple (a, b, c) of integers, with $a, b, c > 0,\ \gcd(a, b) = 1$ and

$$a^2 + b^2 = c^2. \qquad (1)$$

Clearly, if (a, b, c) is a PPT, then one of a, b must be odd, and the other even. Barning [9], and later independently several others [7, 11, 12, 15, 16, 18, 21] (see also [20]), showed:

*Department of Statistics, University of California at Berkeley , 367 Evans Hall, Berkeley CA 94720-3860

Theorem 1. *Define the matrices*

$$M_1 = \begin{pmatrix} -1 & 2 & 2 \\ -2 & 1 & 2 \\ -2 & 2 & 3 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 3 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 1 & -2 & 2 \\ 2 & -1 & 2 \\ 2 & -2 & 3 \end{pmatrix}. \quad (2)$$

Any PPT (a, b, c) *with* a *odd and* b *even has a unique representation as the matrix product*

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = M_{d_1} M_{d_2} \ldots M_{d_n} \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}, \quad (3)$$

for some $n \geq 0$, $(d_1, d_2, \ldots, d_n) \in \{1, 2, 3\}^n$. *Any PPT* (a, b, c) *with* a *even and* b *odd has a unique representation as*

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = M_{d_1} M_{d_2} \ldots M_{d_n} \begin{pmatrix} 4 \\ 3 \\ 5 \end{pmatrix}, \quad (4)$$

for some $n \geq 0$, $(d_1, d_2, \ldots, d_n) \in \{1, 2, 3\}^n$. *Any triple of one of the forms* (3), (4) *is a PPT.*

In some of the papers where this was discussed, the theorem has been described as placing the PPTs (a, b, c) with a odd and b even on the nodes of an infinite rooted ternary tree, with the root representing the "basic" triple $(3, 4, 5)$, and where each triple (a, b, c) has three children, representing the multiplication of the triple (considered as a column vector) by the three matrices M_1, M_2, M_3. In this paper, we consider a slightly different outlook. We think of the sequence $(d_1, d_2, \ldots, d_n)$ in (3), (4) as an *expansion* corresponding to the triple (a, b, c), over the ternary alphabet $\{1, 2, 3\}$. We call the d_i's the *digits* of the expansion. To distinguish between PPTs with a odd, b even and those with a even, b odd, we affix to the expansion a final digit d_{n+1}, which can take the values *oe* (a odd, b even) or *eo* (a even, b odd). So we have a 1-1 correspondence

$$(a, b, c) \in \text{PPT} \quad \longleftrightarrow \quad (d_1, d_2, \ldots, d_{n+1}) \in \bigcup_{n=0}^{\infty} \{1, 2, 3\}^n \times \{oe, eo\}.$$

Several questions now come to mind:

- Is there a simple way to compute the expansion of a PPT? *(Yes – this is contained in the proof of Theorem 1.)*

- As is easy to see and has been known since ancient times, the mapping $(a, b, c) \to (a/c, b/c)$ gives a 1-1 correspondence between the set of PPTs and the rational points (x, y) on the positive quadrant $\mathcal{Q}$ of the unit circle. Are the digits in the expansion piecewise continuous functions on the quadrant? *(Yes.)* Can one define a ternary expansion for *irrational points*? *(Yes.)*

- Can interesting things be said about the expansion $(d_1, d_2, \dots, d_{n+1})$ of a *random* PPT, chosen from some natural model for random PPTs, say by choosing uniformly at random from all PPTs (a, b, c) with $c \le N$ and letting $N \to \infty$? *(Yes.)*

- Do these questions lead to interesting mathematics? *(Yes – they lead to a dynamical system on $\mathcal{Q}$ with interesting properties.)*

It is the goal of this paper to answer these questions. The basic observation is that it is in many ways preferable to deal with points (x, y) on the positive quadrant $\mathcal{Q}$ of the unit circle, instead of with PPTs. In the proof of Theorem 1, we shall see that there is a simple transformation (a piecewise linear mapping) that takes a PPT (a, b, c) to the PPT (a', b', c') that corresponds to its *parent* on the ternary tree, i.e., if (a, b, c) has expansion $(d_1, \dots, d_{n+1})$ then (a', b', c') has expansion $(d_2, \dots, d_{n+1})$. A standard method in dynamical systems is to rescale such transformations, discarding information that is irrelevant for the continuing application of the transformation. For example, the Euclidean algorithm mapping $(x, y) \to (y, x \bmod y)$ is rescaled to the continued fraction mapping $u \to \{1/u\}$ (where $\{z\} = z - \inf\{n \in \mathbb{Z} : n \le z\}$ is the fractional part of a real number z), by replacing the pair of integers (x, y) with the rational number $u = y/x$ and noting that $(x \bmod y)/y = \{1/u\}$. (See also [17, Section 4.5.3], [27].)

When we apply the rescaling idea to our case, we obtain the following result.

Theorem 2. *Let*

$$\mathcal{Q} = \{(x, y) : x > 0,\ y > 0,\quad x^2 + y^2 = 1\}.$$

Define the transformation $T : \mathcal{Q} \to \overline{\mathcal{Q}}$ *by*

$$T(x,y) = \left(\frac{|2-x-2y|}{3-2x-2y}, \frac{|2-2x-y|}{3-2x-2y} \right).$$

Define $d : \mathcal{Q} \to \{1,2,3,oe,eo\}$ *by*

$$d(x,y) = \begin{cases} 1 & 4/3 < x/y, \\ 2 & 3/4 < x/y < 4/3, \\ 3 & x/y < 3/4, \\ & \\ oe & (x,y) = \left(\frac{3}{5}, \frac{4}{5}\right), \\ eo & (x,y) = \left(\frac{4}{5}, \frac{3}{5}\right). \end{cases}$$

Then:

(i) If $(x,y) = (a/c, b/c) \in \mathcal{Q} \cap \mathbb{Q}^2$ *is a rational point of* $\mathcal{Q}$*, with* a/c*,* b/c *in lowest terms (so* (a,b,c) *is a PPT), then for some* $n \geq 0$*,* $T^{n+1}(x,y)$ *(the* $(n+1)$*-th iterate of* T*) will be equal to* $(1,0)$ *or* $(0,1)$*, and if we define*

$$d_k = d(T^{k-1}(x,y)), \qquad k = 1,2,\ldots,n+1,$$

then $(d_1, d_2, \ldots, d_{n+1})$ *is the ternary expansion (with the last digit in* $\{oe, eo\}$*) corresponding to the PPT* (a,b,c) *as in* (3)*,* (4)*.*

(ii) If $(x,y) \in \mathcal{Q}$ *is an irrational point, then* $T^n(x,y) \in \mathcal{Q}$ *for all* $n \geq 0$*, and the sequence*

$$d_k = d(T^{k-1}(x,y)), \qquad k \geq 1,$$

defines an infinite expansion for (x,y) *over the alphabet* $\{1,2,3\}$*, with the property that it does not terminate with an infinite succession of* 1*'s or with an infinite succession of* 3*'s.*

(iii) Any sequence $(d_k)_{k \geq 1}$ *over the alphabet* $\{1,2,3\}$ *which does not terminate with an infinite succession of* 1*'s or an infinite succession of* 3*'s determines a unique (irrational) point* $(x,y) \in \mathcal{Q}$ *such that* $d_k = d(T^{k-1}(x,y)),\quad k \geq 1$*.*

Examples. Here are some examples of points in $\mathcal{Q}$ and their expansions. If $(d_k)_{1\leq k\leq n}$ is an expansion (finite or infinite), we denote by $[d_1, d_2, \ldots]$ the point $(x,y)\in\mathcal{Q}$ which has the given sequence as its expansion.

$$\begin{aligned}(3/5,4/5) &= [oe]\\ (4/5,3/5) &= [eo]\end{aligned} \qquad \begin{aligned}(15/17,8/17) &= [1,oe]\\ (21/29,20/29) &= [2,oe]\\ (5/13,12/13) &= [3,oe]\end{aligned}$$

$$\begin{aligned}(35/37,12/37) &= [1,1,oe] & (65/97,72/97) &= [2,1,oe] & (33/65,56/65) &= [3,1,oe]\\ (77/85,36/85) &= [1,2,oe] & (119/169,120/169) &= [2,2,oe] & (39/89,80/89) &= [3,2,oe]\\ (45/53,28/53) &= [1,3,oe] & (55/73,48/73) &= [2,3,oe] & (7/25,24/25) &= [3,3,oe]\end{aligned}$$

$$\begin{aligned}(\sqrt{2}/2,\sqrt{2}/2) &= [2,2,2,2,\ldots] & (2/\sqrt{5},1/\sqrt{5}) &= [1,2,1,2,\ldots]\\ (1/2,\sqrt{3}/2) &= [3,1,3,1,\ldots] & (1/\sqrt{5},2/\sqrt{5}) &= [3,2,3,2,\ldots]\\ (\sqrt{3}/2,1/2) &= [1,3,1,3,\ldots] & (3/\sqrt{10},1/\sqrt{10}) &= [1,1,2,1,1,2,\ldots]\end{aligned}$$

$$(\cos 1,\sin 1) = [3,1,1,3,1,1,1,1,3,1,1,1,1,1,1,3,1,1,1,1,1,1,1,1,3,\ldots]$$

(see section 5)

$$(\cos(1/\pi),\sin(1/\pi)) = [1,1,2,1,2,2,3,3,3,3,3,3,3,1,3,3,3,3,3,3,2,\ldots]$$

(this is meant as an example of a "typical" expansion – see section 4)

$$[1,1,\ldots,1,oe] \ \ (n \text{ times "1"}) = \left(\frac{4(n+1)^2-1}{4(n+1)^2+1}, \frac{4(n+1)}{4(n+1)^2+1}\right)$$

$$[2,2,\ldots,2,oe] \ \ (n \text{ times "2"}) = \left(\frac{a_n}{c_n}, \frac{a_n+(-1)^n}{c_n}\right)$$

where $(a_n)_{n\geq 0} = \left(\frac{(\sqrt{2}+1)^{2n+3}-(\sqrt{2}-1)^{2n+3}-2(-1)^n}{4}\right)_{n\geq 0} = (3,21,119,697,\ldots)$ and $(c_n)_{n\geq 0} = \left(\frac{(2-\sqrt{2})(3-2\sqrt{2})^{n+1}+(2+\sqrt{2})(3-2\sqrt{2})^{n+1}}{4}\right)_{n\geq 0} = (5,29,169,985,\ldots)$ are sequences A046727 and A001653, respectively, in The On-Line Encyclopedia of Integer Sequences [26]. The triples $(a_n, a_n+(-1)^n, c_n)$ are the integer solutions to the equation $x^2+(x+1)^2=z^2$ – that is, the best approximations in integers to the 45-45-90 triangle; c_n is the n-th integer such that $2n^2-1$ is a square.

After constructing the dynamical system associated with the ternary expansion, the next step is to study its properties. What does the expansion of a typical point look like? The answer is essentially contained in the following theorem.

Theorem 3. *Let ds denote arc length on the unit circle. The dynamical system $(\mathcal{Q}, T)$ possesses an infinite invariant measure μ, given by*

$$d\mu(x,y) = \frac{ds}{\sqrt{(1-x)(1-y)}}. \tag{5}$$

With the measure μ, the system $(\mathcal{Q}, T, \mu)$ is a conservative and ergodic infinite measure-preserving system.

The invariant measure μ encodes all the information about the statistical regularity of expansions of "typical" points. In section 4 we shall state more explicitly some of the number-theoretic consequences of Theorem 3.

Recall ([13, Theorem 225]) that the general parametric solution of the equation (1) with $a, b > 0$ coprime, a odd and b even is given by

$$a = m^2 - n^2, \qquad b = 2mn, \qquad c = m^2 + n^2, \tag{6}$$

where m, n have opposite parity, $m > n > 0$, and $\gcd(m, n) = 1$. A roughly equivalent statement is that the map

$$D : t \longrightarrow \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2} \right)$$

maps the extended real line injectively onto the unit circle, and maps the rational numbers, together with the point at infinity, onto the rational points of the circle. Note also that the interval $(0, 1)$ is mapped onto the positive quadrant $\mathcal{Q}$. It is thus natural, when trying to understand the behavior of the dynamical system $(\mathcal{Q}, T)$, to conjugate it by the mapping D, to obtain a new dynamical system on $(0, 1)$. This leads to the following result (the precise meaning of the last statement in the theorem will be explained later):

Theorem 4. *$(\mathcal{Q}, T, \mu)$ is conjugate, by the mapping D, to the measure preserving system $((0,1), \hat{T}, \nu)$, where*

$$\begin{aligned} \hat{T}(t) &= (D^{-1} \circ T \circ D)(t) = \begin{cases} \frac{t}{1-2t} & 0 < t < \frac{1}{3}, \\ \frac{1}{t} - 2 & \frac{1}{3} < t < \frac{1}{2}, \\ 2 - \frac{1}{t} & \frac{1}{2} < t < 1, \end{cases} \\ d\nu(t) &= \frac{1}{\sqrt{2}} \cdot \frac{dt}{t(1-t)}. \end{aligned}$$

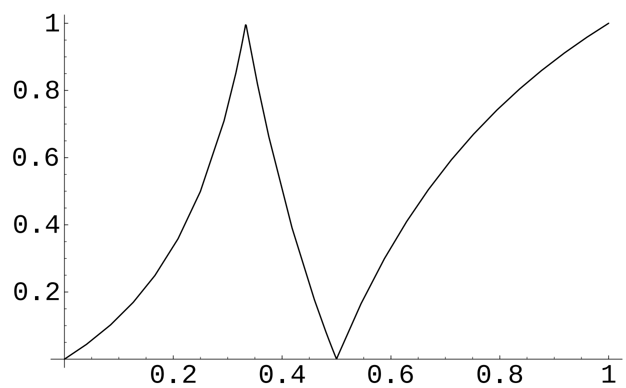


Figure 1: The interval map $\hat{T}$

The mapping $\hat{T}$ is the scaling of a modified slow (subtractive) Euclidean algorithm.

There has been some interest in obtaining natural dynamical systems appearing in number theory as factors of certain cross sections of the geodesic flow on quotients of the hyperbolic plane by a discrete subgroup of its isometry group. This has been done by Adler and Flatto [4, 6] and by Series [24] for the continued fraction transformation, and by Adler and Flatto [5] for Rényi's backward continued fraction map (see also [3, 25]). In both of these cases, the underlying surface was the modular surface, which is the quotient of the hyperbolic plane by the modular group $\Gamma = PSL(2, \mathbb{Z})$. Representing a system as a factor of a cross-section of a geodesic flow enables one to derive mechanically (without guessing) an expression for the invariant density, and to deduce various properties of the system. Series ([25], Problem 5.25(i)) posed the general problem of replicating this idea for other number- and group-theoretical dynamical systems.

It turns out that the map T also admits such a representation. Aaronson and Denker [2] studied a certain cross-section of the geodesic flow on a different quotient of the hyperbolic plane, which is the quotient by the congruence subgroup $\Gamma(2)$ of the modular group, a free group with two generators. They obtained the map $\hat{T}$ as a factor of that cross section, and used this to derive results on the asymptotic behavior of the Poincaré series of the group $\Gamma(\mathbb{C}\setminus\mathbb{Z})$ of deck transformations of $\mathbb{C}\setminus\mathbb{Z}$. Since in their paper the motivation was completely different from ours, and the connection to the map T was not known, we find it worthwhile to include here a version of their result.

Theorem 5. *Let $(\mathbb{H}, (\varphi_t)_{t\in\mathbb{R}})$ be the upper half-plane model of the hyperbolic plane, with the associated geodesic flow $\varphi_t : T_1(\mathbb{H}) \to T_1(\mathbb{H})$, where $T_1(\mathbb{H}) = \mathbb{H} \times S^1$ is the unit tangent bundle of $\mathbb{H}$. Define the group of isometries of $\mathbb{H}$*

$$\Gamma(2) = \left\{ z \to \frac{az+b}{cz+d} : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}),\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ mod\ 2 \right\},$$

and let $M = \mathbb{H}/\Gamma(2)$ be the quotient space of $\mathbb{H}$ by $\Gamma(2)$, which has a fundamental domain

$$F = \left\{ z \in \mathbb{H} : |Re\ z| < 1,\ \left| z \pm \frac{1}{2} \right| > \frac{1}{2} \right\}.$$

(The surface M is the unique complete hyperbolic structure on the 3-punctured sphere.) Let $\pi : \mathbb{H} \to M$ be the quotient map. Let $(\overline{\varphi}_t)_{t\in\mathbb{R}}$ be the geodesic flow on M. Let $X' \subset T_1(\mathbb{H})$,

$$X' = \left\{ (z,u) \in \partial F \times S^1 : z + \epsilon u \in F \ for\ small\ \epsilon > 0 \right\},$$

and let $X \subset T_1(M)$, $X = d\pi(X')$ be the natural section of M corresponding to the fundamental domain F of all inward-pointing vectors on the boundary of F. Let $\tau : X \to X$ be the section- or first-return map of the geodesic flow, namely

$$\tau(\omega) = \overline{\varphi}_{t_\omega}(\omega),$$

where

$$t_\omega = \inf\{t > 0 : \overline{\varphi}_t(\omega) \in X\}.$$

Then the section map (X, τ) admits the map $(\mathcal{Q}, T)$ as a factor. That is, there exists an (explicit) function $E : X \to \mathcal{Q}$ such that $T \circ E = E \circ \tau$.

Theorem 5 is an immediate consequence of Aaronson and Denker's result ([2, Section 4]) and Theorem 4. To describe explicitly the factor map E, define $p_1(x, \delta, \epsilon) = x$. Then, in the notation of their paper,

$$E = D \circ p_1 \circ \eta^{-1} \circ \pi_+,$$

with the "juicy" parts being our map D and and the map π_+, which assigns to a tangent vector the hitting point on the real axis of the geodesic emanating from the lifting of the tangent vector to $\partial F \times S^1$. For more details consult [2]. Figure 2 shows the tesselation of the hyperbolic plane by $\Gamma(2)$-translates of the fundamental domain F, shown in the Poincaré disk model of the hyperbolic plane.

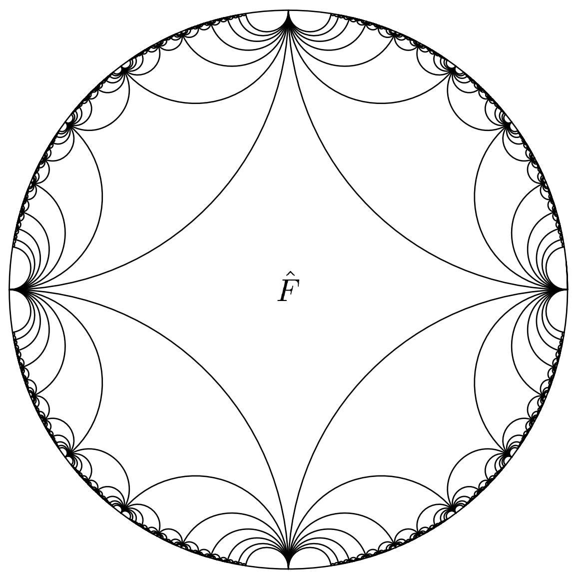


Figure 2: The $(\hat{\Gamma}(2), \hat{F})$-tesselation of the Poincaré disk $\mathbb{D}$, where $\hat{\Gamma}(2)$ is the group generated (freely) by the Möbius transformations $z \to \frac{(1-i)z-1}{z-1-i}$, $z \to \frac{(1+i)z+1}{-z-1+i}$ and $\hat{F} = \{z \,:\, |z - (\pm 1 \pm i)| > 1\}$.

The congruence subgroup $\Gamma(2)$ also appears in the paper by Alperin [7], which discusses the ternary tree of PPTs.

In the next section, we reprove Theorem 1, and show how the linear mappings involved in the construction of the ternary tree of PPTs can be scaled down to produce the transformation T. This will result in a proof of Theorem 2. In section 3, we prove Theorem 4 and discuss the connection to modified Euclidean algorithms. The ergodic properties of the system will be derived, using standard techniques of infinite ergodic theory, proving Theorem 3. In section 4 we discuss applications to the statistics of expansions of random points on $\mathcal{Q}$ and random PPTs. Section 5 has some remarks on points with special expansions and possible directions for further study.

2 Construction of the dynamical system

2.1 The piecewise linear transformation

First, we recall the ideas involved in the proof of Theorem 1. We follow the elegant exposition of [19].

We shall consider solutions of (1) with $\gcd(a,b)=1$, and $c>0$. Define

$$\begin{aligned} \mathrm{PPT} &= \{(a,b,c)\in\mathbb{Z}^3 : \gcd(a,b)=1,\ \ a,b,c>0,\ a^2+b^2=c^2\}, \\ &\qquad \text{the set of PPTs, and} \\ \mathrm{SPPT} &= \{(a,b,c)\in\mathbb{Z}^3 : \gcd(a,b)=1,\ \ c>0,\ a^2+b^2=c^2\}, \\ &\qquad \text{the set of } \textit{signed} \text{ PPTs.} \end{aligned}$$

The basic observation is that the equation (1) has three symmetries. Two of them are the obvious symmetries $a\to -a$, $b\to -b$ (we ignore the symmetry $c\to -c$, since we are only considering solutions with $c>0$). The third symmetry is not so obvious, but becomes obvious when the correct change of variables is applied. Define new variables m,n,q by

$$\begin{aligned} m &= c-a \\ n &= c-b \\ q &= a+b-c \end{aligned} \qquad \longleftrightarrow \qquad \begin{aligned} a &= q+m \\ b &= q+n \\ c &= q+m+n \end{aligned}$$

In the new variables, (1) becomes

$$q^2 = 2mn. \tag{7}$$

There is therefore a third natural involution on the set SPPT of solutions of (1), given in m,n,q coordinates by $q\to -q$. So, starting from a solution $(a,b,c)\in\mathrm{SPPT}$ with associated variables (m,n,q) and setting $q'=-q, m'=m, n'=n$, we arrive at a new solution (a',b',c') given by

$$\begin{aligned} a' &= q'+m' &&= a-2q &&= 2c-a-2b \\ b' &= q'+n' &&= b-2q &&= 2c-2a-b \\ c' &= q'+m'+n' &&= c-2q &&= 3c-2a-2b, \end{aligned}$$

or in matrix notation

$$\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} -1 & -2 & 2 \\ -2 & -1 & 2 \\ -2 & -2 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} =: I \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

The matrix I is an involution, i.e. $I^2 = id_3$. It is easy to see that a', b' are also coprime, and

$$c' = 3c - 2(a+b) \geq 3c - 2\sqrt{2} \cdot \sqrt{a^2+b^2} = (3 - 2\sqrt{2})c > 0.$$

So I maps SPPT to itself. I has the fixed points $(1,0,1)$ and $(0,1,1)$. We claim that $I(\text{PPT}) = \text{SPPT} \setminus (\text{PPT} \cup \{(1,0,1),(0,1,1)\})$. Indeed, this simply means that if $(a,b,c) \in \text{PPT}$, then at least one of a', b' is negative, or in other words that $2c < \max(a+2b, 2a+b)$. Assume for concreteness that $a > b$, then $2c < 2a + b$ if $2 < 2x + y$, where $(x,y) = (a/c, b/c) \in \mathcal{Q} \cap \{x > y\}$, or equivalently if

$$\frac{2}{\sqrt{5}} < \langle\ (x,y),\ (2/\sqrt{5}, 1/\sqrt{5})\ \rangle. \tag{8}$$

The point $(2/\sqrt{5}, 1/\sqrt{5})$ lies on the arc $\mathcal{Q} \cap \{x > y\}$, and one checks easily that there is an equality at one end $(1,0)$ of the arc, and a strict inequality at the other end $(1/\sqrt{2}, 1/\sqrt{2})$. So (8) holds.

Having shown that if $(a,b,c) \in \text{PPT}$, then (a',b',c') is a signed PPT with one of a, b negative, we can forget about the signs of a', b' and obtain a new triple $(a'',b'',c'') = (|a'|,|b'|,c')$. c' is strictly less than c, since $c' = c - 2q$ and $q = a + b - c = a + b - \sqrt{a^2+b^2} > 0$ on PPT. The new triple will be a PPT, except when $(a,b,c) = (3,4,5)$ or $(4,3,5)$, in which case (a'',b'',c'') will equal $(1,0,1)$ or $(0,1,1)$, respectively. If (a'',b'',c'') is a PPT, there are precisely three PPTs (a,b,c) leading to it via this procedure – corresponding to the three possible sign patterns $a' < 0 < b'$; $a', b' < 0$; $a' > 0 > b'$ (we ruled out $a', b' > 0$) – and they can easily be recovered, as follows: If $a' < 0 < b'$, then

$$\begin{pmatrix} a'' \\ b'' \\ c'' \end{pmatrix} = \begin{pmatrix} -a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} I \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$
$$\Longrightarrow \begin{pmatrix} a \\ b \\ c \end{pmatrix} = I \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a'' \\ b'' \\ c'' \end{pmatrix} = M_1 \begin{pmatrix} a'' \\ b'' \\ c'' \end{pmatrix}.$$

(with M_1 as in (2)). Similarly we get

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = M_2 \begin{pmatrix} a'' \\ b'' \\ c'' \end{pmatrix}, \text{ if } a', b' < 0, \text{ or } \begin{pmatrix} a \\ b \\ c \end{pmatrix} = M_3 \begin{pmatrix} a'' \\ b'' \\ c'' \end{pmatrix}, \text{ if } a' > 0 > b'.$$

We are now ready to prove Theorem 1. First, from the above discussion it follows that M_1, M_2, M_3 take PPTs to PPTs with a strictly larger third coordinate. In particular, any triple of one of the forms (3), (4) is a PPT. Next, for $(a,b,c) \in \mathrm{PPT}$ define

$$
\begin{aligned}
S(a,b,c) &= (|2c-a-2b|, |2c-2a-b|, 3c-2a-2b), \\
\delta(a,b,c) &= \begin{cases} 1 & 2c-a-2b < 0 < 2c-2a-b \\ 2 & 2c-a-2b, 2c-2a-b < 0 \\ 3 & 2c-a-2b > 0 > 2c-2a-b \end{cases}, \\
\delta_k(a,b,c) &= \delta(S^{k-1}(a,b,c)), \quad k = 1,2,\ldots,n(a,b,c), \\
n(a,b,c) &= \max\{n \geq 0 : S^n(a,b,c) \in \mathrm{PPT}\}.
\end{aligned}
$$

The above discussion can be summarized by the equations

$$
\begin{pmatrix} a \\ b \\ c \end{pmatrix} = M_{\delta(a,b,c)}(S(a,b,c))^{\mathrm{t}}, \qquad \delta\left(M_d \begin{pmatrix} a \\ b \\ c \end{pmatrix}\right) = d \quad (d = 1,2,3). \tag{9}
$$

Let $(a,b,c) \in \mathrm{PPT}$ with a odd and b even. It is easy to see that M_1, M_2, M_3 preserve the parity of a, b, so (a,b,c) cannot have a representation (4). We claim that it satisfies (3) with $d_k = \delta_k(a,b,c)$, $k = 1,\ldots,n(a,b,c)$, and that this representation is unique. The proof is by induction on c. The claim holds for the basic triple $(3,4,5)$, because M_1, M_2, M_3 increase the third coordinate. Assume that it holds for all odd-even PPTs with third coordinate $< c$. Then in particular this is true for $(a',b',c') = S(a,b,c)$, since we know that $c' < c$. So we may write

$$
\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = M_{e_1} M_{e_2} \ldots M_{e_m} \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}
$$

with $e_k = \delta_k(a',b',c')$, $1 \leq k \leq m = n(a',b',c') = n(a,b,c) - 1$. We have

$$
e_k = \delta(S^{k-1}(a',b',c')) = \delta(S^k(a,b,c)) = \delta_{k+1}(a,b,c) = d_{k+1},
$$

where we denote $d_k = \delta_k(a,b,c)$. Therefore, by (9),

$$
\begin{pmatrix} a \\ b \\ c \end{pmatrix} = M_{\delta(a,b,c)} \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = M_{d_1} M_{d_2} \ldots M_{d_{m+1}} \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix},
$$

which is our claimed representation. Uniqueness follows immediately by noting that by (9), d_1 is determined by the sign pattern of $(2c-a-2b, 2c-2a-b)$, and continuing by induction. Theorem 1 is proved.

2.2 Scaling the transformation

It is now easy to rescale the transformation S to obtain a transformation T from $\mathcal{Q}$ to its closure. If $(x,y)=(a/c,b/c)\in\mathcal{Q}\cap\mathbb{Q}^2$ is a rational point of $\mathcal{Q}$, which corresponds to the PPT (a,b,c), then the triple $S(a,b,c)=(|2c-a-2b|,|2c-2a-b|,3c-2a-2b)$ corresponds to the point

$$\left(\frac{|2c-a-2b|}{3c-2a-2b},\frac{|2c-2a-b|}{3c-2a-2b}\right)=\left(\frac{|2-x-2y|}{3-2x-2y},\frac{|2-2x-y|}{3-2x-2y}\right)$$

in $\overline{\mathcal{Q}}$. This precisely accounts for our definition of T in Theorem 2. To explain why the function d is the correct rescaling of δ, observe, for example, that $\delta(a,b,c)=1$ iff

$$2-x-2y<0<2-2x-y \iff \begin{array}{c}\frac{2}{\sqrt{5}}<\langle\,(x,y),(1/\sqrt{5},2/\sqrt{5})\,\rangle\\ \langle\,(x,y),(2/\sqrt{5},1/\sqrt{5})\,\rangle<\frac{2}{\sqrt{5}},\end{array}$$

which some inspection reveals to hold exactly on the (open) circular arc connecting the point $(1,0)$ with the point $(4/5,3/5)$. This corresponds to the condition $x/y>4/3$ in the definition of d. Similarly, it can be checked that $\delta(a,b,c)=2$ if (x,y) lies on the circular arc between $(4/5,3/5)$ and $(3/5,4/5)$, and $\delta(a,b,c)=3$ if (x,y) lies on the circular arc between $(3/5,4/5)$ and $(0,1)$. These arcs form the *generating partition* of the ternary expansion – see Figure 3.

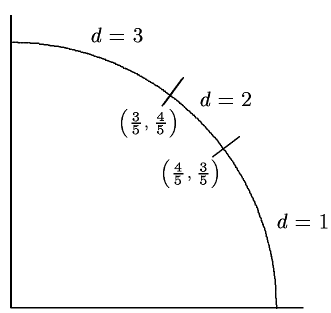


Figure 3: The quadrant $\mathcal{Q}$ and the generating partition

Figure 4 shows the graph of the map obtained by parametrizing the quadrant $\mathcal{Q}$ in terms of the angle (multiplied by $2/\pi$, to obtain a map on the

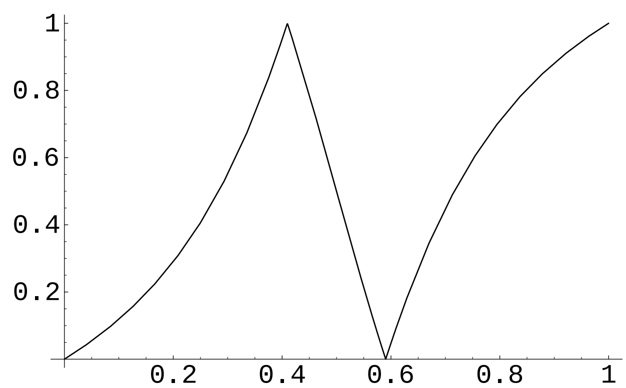


Figure 4: The conjugate map $F^{-1} \circ T \circ F$, where $F(t) = (\cos(\pi t/2), \sin(\pi t/2))$

interval $(0,1)$). As Theorem 4 may imply, this is not the best parametrization, but it gives a good graphical illustration of the behavior of the mapping T. Note that, contrary to appearance, the map is not linear on the middle interval!

We have proved part (i) of Theorem 2. For part (ii), observe that all the iterates of an irrational point $(x,y) \in \mathcal{Q}$ under T remain irrational, since T is defined by piecewise rational functions with integer coefficients which are invertible (so as before, given $d(x,y)$, we can recover (x,y) from $T(x,y)$ by a rational function with integer coefficients, hence if $T(x,y)$ is rational, so is (x,y)). For $T^n(x,y)$ to be in $\overline{\mathcal{Q}} \setminus \mathcal{Q} = \{(1,0),(0,1)\}$, we must have $T^{n-1}(x,y) = (3/5, 4/5)$ or $(4/5, 3/5)$, and this cannot happen for an irrational point. Therefore $T^n(x,y)$ is defined for all $n \geq 1$, as claimed. The resulting sequence of digits $d_k = d(T^{k-1}(x,y))$ cannot terminate with an infinite succession of 1's. Indeed, on the first interval $(0, \frac{2}{\pi}\arctan(3/4))$ of the generating partition of the mapping $G = F^{-1} \circ T \circ F$ (Figure 4) it is easy to verify that G' is increasing and satisfies $G'(0) = 1$. Therefore if t is some point on this interval, then the sequence of iterates $G^k(t)$ satisfies $G^k(t) \geq G'(t)^k \cdot t$ and therefore must eventually leave this interval (so the corresponding succession of 1's in the expansion will terminate). A symmetrical argument applies for the third interval of the generating partition, implying that no infinite expansion terminates with an infinite succession of 3's, and part (ii) is proved.

Turn to the final part (iii) of Theorem 2. Again we use the mapping G

in Figure 4. Let $I_1 = (0, \frac{2}{\pi}\arctan(3/4))$, $I_2 = (\frac{2}{\pi}\arctan(3/4), \frac{2}{\pi}\arctan(4/3))$, $I_3 = (\frac{2}{\pi}\arctan(4/3), 1)$ be the intervals of the generating partition of G. Given an infinite expansion $(d_k)_{k\geq 1}$ that does not terminate with an infinite succession of 1's or an infinite succession of 3's, we must show that there is a unique number $t \in (0,1)$ such that $G^{k-1}(t) \in I_{d_k}$ for all $k \geq 1$.

Consider, for any $n \geq 1$, the *cylinder set*

$$A_n = \{t \in (0,1) : G^{k-1}(t) \in I_{d_k} \text{ for } 1 \leq k \leq n\}.$$

$(A_n)_{n\geq 1}$ is a decreasing sequence of non-empty open intervals. By compactness, the intersection of their closures contains at least one point t. The condition on the sequence $(d_k)_{k\geq 1}$ implies that t is in fact in the intersection of the *open* intervals; otherwise, t is an endpoint, say of A_n, but that would imply that $d_k = 1$ for all $k > n$ or $d_k = 3$ for all $k > n$.

We have shown existence of a number with a prescribed expansion. But uniqueness also follows, since, as Figure 4 shows, any appearance of a "2" digit, or a non-1 digit following a succession of 1's, or a non-3 digit following a succession of 3's, entails a shrinkage of the corresponding set A_n by at least a constant factor bounded away from 1. So the diameter of the A_n's goes to 0, and their intersection contains at most one point. Theorem 2 is proved.

(Here is another argument demonstrating uniqueness: any two irrational points on $\mathcal{Q}$ are separated by a rational point; after a finite number of applications of T, the rational point will be mapped to $(3/5, 4/5)$ or to $(4/5, 3/5)$, and the images of the two irrational points will be contained in different elements of the generating partition, implying a different first digit in their expansions.)

3 The modified Euclidean algorithm

3.1 Some computations

The inverse function of D is easily computed to be

$$D^{-1}(x,y) = \frac{1-x}{y}.$$

Using this, a routine computation, which we omit, shows that indeed

$$\hat{T} = (D^{-1} \circ T \circ D).$$

We show that the measure ν is $\hat{T}$-invariant. If $d\nu(t) = f(t)dt$, the invariant density must satisfy

$$f(t) = \sum_{u=\hat{T}^{-1}(t)} f(u) \cdot \frac{1}{|\hat{T}'(u)|}. \tag{10}$$

The inverse branches of $\hat{T}$ are given by

$$\begin{aligned} F_1(t) &= \frac{t}{1+2t} \in (0, 1/3), \\ F_2(t) &= \frac{1}{2+t} \in (1/3, 1/2), \\ F_3(t) &= \frac{1}{2-t} \in (1/2, 1), \end{aligned} \tag{11}$$

for which

$$\begin{aligned} |T'(F_1(t))| &= (1 - 2F_1(t))^{-2} = (1+2t)^2, \\ |T'(F_2(t))| &= F_2(t)^{-2} = (2+t)^2, \\ |T'(F_3(t))| &= F_3(t)^{-2} = (2-t)^2. \end{aligned}$$

So (10) reduces to

$$f(t) = f\left(\frac{t}{1+t}\right)\frac{1}{(1+2t)^2} + f\left(\frac{1}{2+t}\right)\frac{1}{(2+t)^2} + f\left(\frac{1}{2-t}\right)\frac{1}{(2-t)^2}. \tag{12}$$

Check directly that $f(t) = 1/(t(1-t))$ satisfies (12).

To complete the proof of Theorem 4, we need to verify that μ is the push-forward of the measure ν under D. Denote

$$x = x(t) = \frac{1-t^2}{1+t^2}, \qquad y = y(t) = \frac{2t}{1+t^2}.$$

Compute:

$$\frac{1}{\sqrt{(1-x)(1-y)}} = \left(\frac{2t^2}{1+t^2} \cdot \frac{(1-t)^2}{1+t^2}\right)^{-1/2} = \frac{1+t^2}{\sqrt{2}\, t(1-t)}.$$

$$\begin{aligned} ds &= \sqrt{dx^2+dy^2} = \sqrt{x'(t)^2+y'(t)^2}dt \\ &= \sqrt{\left(\frac{-4t}{(1+t^2)^2}\right)^2 + \left(\frac{2(1-t^2)}{(1+t^2)^2}\right)^2} dt = \frac{2\,dt}{1+t^2}. \\ &\implies \frac{dt}{\sqrt{2}\,t(1-t)} = \frac{ds}{\sqrt{(1-x)(1-y)}}, \end{aligned}$$

as claimed.

Note that this also proves that μ is T-invariant. This fact could be checked directly, of course.

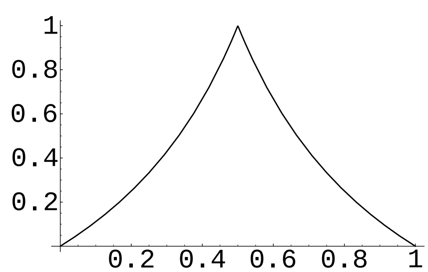
3.2 Interpretation as a Euclidean algorithm

The ordinary Euclidean algorithm takes a pair of positive integers (x, y) with $x > y$ and returns the pair $(y, x \bmod y)$. After a finite number of iterations of this mapping, y will be equal to 0 and x will be equal to the g.c.d. of the original pair.

Many variants of this algorithm have been analyzed, where various alternatives to simple division with remainder are used. The study of such algorithms, related of course to continued fraction variants, is a huge subject which it is beyond the scope of this paper to describe. See [23]; sections 4.5.2-4.5.3 in [17] and the references there; and [8, 27] for some more recent developments.

The standard Euclidean algorithm has a more ancient version, known as the *slow*, or *subtractive* Euclidean algorithm, where subtraction is used instead of division, so (x, y) are mapped to $(\max(x-y, y), \min(x-y, y))$. Clearly the standard algorithm is nothing more than a speeding-up of this algorithm. One may scale by always replacing x by 1 and y by the ratio y/x. This leads to the interval map $R : (0, 1) \to (0, 1)$,

$$R(t) = \begin{cases} \frac{t}{1-t} & 0 < t \le 1/2, \\ \frac{1-t}{t} & 1/2 < t < 1 \end{cases} =$$



We now observe that the map $\hat{T}$ is itself the scaling of a modified algorithm, defined by the mapping

$$(x,y),\quad x>y \quad\longrightarrow\quad (a,b)=\begin{cases}(x-2y,y) & \text{if } x-2y>y,\\ (y,x-2y) & \text{if } y\geq x-2y>0,\\ (y,2y-x) & \text{if } x-2y\leq 0.\end{cases}$$

Here is a sample execution sequence of this algorithm:

$$(155,100)\longrightarrow(100,45)\longrightarrow(45,10)\longrightarrow(25,10)\longrightarrow(10,5)\longrightarrow(5,0).$$

This algorithm can be used to compute g.c.d.'s, just like its famous kin: The last output which differs from the one preceding it is of either the form (a,a) or $(a,0)$, where a is the g.c.d. of the two original integers. This is easy to prove by induction using the observation that each of the three operations performed by the algorithm preserves g.c.d.'s and that if $x>y>0$ then $\max(a,b)<\max(x,y)$.

3.3 Ergodic properties of $\hat{T}$

To prove Theorem 3, we study the somewhat simpler measure preserving system $((0,1),\hat{T},\nu)$. Since this is now represented as an interval map, we can use standard techniques of ergodic theory.

Define $\hat{d}(t)=d(D(t))$. Define the set $J=(1/5,2/3)$. An alternative description for J is as the cylinder set

$$J=\Big\{t\in(0,1):\hat{d}(t)=2 \text{ or } \big(\hat{d}(t)=1 \text{ and } \hat{d}(\hat{T}(t))\neq 1\big)$$
$$\text{or } \big(\hat{d}(t)=3 \text{ and } \hat{d}(\hat{T}(t))\neq 3\big)\Big\}.$$

By Theorem 2(ii), for any irrational $t\in(0,1)$, $\hat{T}^n(t)\in J$ for some $n\geq 0$. In other words,

$$(0,1)=\bigcup_{n=0}^{\infty}\hat{T}^{-n}(J)\quad\text{a.e.}$$

This implies that $\hat{T}$ is conservative, by [1], Theorem 1.1.7.

To prove that $\hat{T}$ is ergodic, we pass to the *induced system* $(J, \hat{T}_J, \mu_{|J})$, where

$$\begin{aligned} \hat{T}_J(t) &= \hat{T}^{\varphi_J(t)}(t), \\ \varphi_J(t) &= \inf\{n \geq 1 : T^n(t) \in J\}. \end{aligned}$$

In [2, p. 16], Aaronson and Denker construct an explicit Markov partition for $\hat{T}_J$, and deduce using standard techniques ([2, Lemma 5.2]) that $\hat{T}_J : J \to J$ is a topologically mixing Markov map which is uniformly expanding with bounded distortion, i.e., satisfies

$$\inf_{t \in J} |\hat{T}_J'(t)| > 1, \qquad \sup_{t \in J} \frac{|\hat{T}_J''(t)|}{(\hat{T}_J'(t))^2} < \infty.$$

Therefore ([1, Theorem 4.4.7]) it is exact, and in particular it is ergodic. It follows ([1, Proposition 1.5.2(2)]) that $\hat{T}$ is itself ergodic. This completes the proof of Theorem 3.

4 Expansions of random $\mathcal{Q}$-points and random PPTs

4.1 Random points on $\mathcal{Q}$

The invariant measure μ becomes infinite near the two ends of the quadrant $\mathcal{Q}$. This means that in a typical expansion, the digits "1" and "3" will occur infinitely more often than the middle digit "2". However, for any two digit sequences, even ones that contain the digits "1" and "3", we can ask about their relative density of occurence in the expansion of typical points.

Theorem 6. *Let $I_1 = (0, 1/3), I_2 = (1/3, 1/2), I_3 = (1/2, 1)$ be the intervals of the generating partition of $\hat{T}$. For $(d_1, \ldots, d_n) \in \cup_{\ell=1}^{\infty} \{1, 2, 3\}^{\ell}$, define*

$$A(d_1, \ldots, d_n) = \nu\left(\bigcap_{j=1}^{n} \hat{T}^{-j+1}(I_{d_j}) \right) = \nu\Big(\big(F_{d_1} \circ F_{d_2} \circ \ldots \circ F_{d_n}\big)\big((0,1)\big) \Big),$$

with F_1, F_2, F_3 as in (11). Let $(d_1, \dots, d_n), (e_1, \dots, e_m) \in \cup_{\ell=1}^{\infty}\{1,2,3\}^{\ell}$. For almost every $(x,y) \in \mathcal{Q}$, the limit

$$\lim_{N\to\infty} \frac{\#\left\{0 \le k \le N : d(T^{k+j-1}(x,y)) = d_j, \quad 1 \le j \le n\right\}}{\#\left\{0 \le k \le N : d(T^{k+j-1}(x,y)) = e_j, \quad 1 \le j \le m\right\}}$$

exists and is equal to $A(d_1, \dots, d_n)/A(e_1, \dots, e_m)$.

Proof. This is an immediate consequence of Hopf's ergodic theorem, applied to the two indicator functions of the cylinder sets $\cap_{j=1}^{n}\hat{T}^{-j+1}(I_{d_j})$ and $\cap_{j=1}^{m}\hat{T}^{-j+1}(I_{e_j})$. ∎

Example. An easy computation gives

$$\frac{A(1,2)}{A(1,3)} = \frac{\log(4/3)}{\log(3/2)} \approx \frac{0.2876}{0.4055}.$$

Therefore, in a typical expansion, when a run of consecutive 1's breaks, the next digit will be a "2" with probability $0.2876/(0.2876 + 0.4055) \approx 0.415$, or a "3" with probability $0.4055/(0.2876 + 0.4055) \approx 0.585$.

4.2 Random PPTs

PPTs have finite expansions and form a subset of $\mathcal{Q}$ of measure 0. So, as the analogous studies of continued fraction expansions of rational numbers (a.k.a. analysis of the Euclidean algorithm) have shown, analyzing their behavior can be significantly more difficult than the behavior of expansions of random points on $\mathcal{Q}$. We outline here a technique for easily deducing some of the properties of the expansion by relating the discrete model to the continuous one. We mention some open problems which may be approachable using more sophisticated methods such as those used in [8], and which we hope to address at a later date.

Our model for random PPTs will be the discrete probability space

$$\mathrm{PPT}_N = \{(a,b,c) \in \mathrm{PPT} : c \le N\},$$

equipped with the uniform probability measure $\mathbb{P}_N$. Analogous results can easily be formulated, using the same ideas presented here, for other natural models, e.g., a uniform choice of $(a, b, c) \in PPT$ with $|a|, |b| \leq N$.

We discuss the distribution of the individual digits in the expansion. Let λ be the uniform arc-length measure on $\mathcal{Q}$, normalized as a probability measure. That is, $d\lambda = (2/\pi)ds$. We need the following simple lemma.

Lemma 7. *Under the measure $\mathbb{P}_N$, the random vector $(a/c, b/c)$ converges in distribution to λ.*

Proof. We prove this for the uniform measure $\mathbb{P}_N^{oe}$ on PPTs $(a, b, c) \in \text{PPT}_N$ with a odd and b even; the claim will follow by symmetry.

Observe the following fact from elementary number theory: the coprime pairs (m, n) such that m, n are of opposite parity have a local density of $4/\pi^2$ in the lattice $\mathbb{Z}^2$, in the following sense: for any bounded open set $D \subset \mathbb{R}^2$, we have

$$\frac{1}{x^2}\#\left\{(m,n) \in \mathbb{Z}^2 : \left(\frac{m}{x}, \frac{n}{x}\right) \in D,\ \gcd(m,n) = 1, \quad m + n \equiv 1 (\text{mod } 2)\right\} \xrightarrow[x\to\infty]{} \frac{4}{\pi^2}\text{area}(D). \tag{13}$$

(This is related to the well-known fact that the coprime pairs, or "lattice points visible from the origin", have density $6/\pi^2$ in the lattice $\mathbb{Z}^2$, see [13, Th. 332, p. 269], [14]. In effect we are claiming that among these coprime pairs, the density of points with unequal parity is $2/3$ – this makes sense since the parities are both random but conditioned not to both be even.)

To prove (13), first we prove that this is true when D is a rectangle $(0, A) \times (0, B)$. Define for $i = 0, 1$,

$$g_i(u; k) = \#\left\{j \in \mathbb{Z} : 0 < j < u, \quad j \equiv i (\text{mod } 2), \quad k \mid j\right\}.$$

Let $(\mu_2(k))_{k\geq 1}$ be the coefficients of the Dirichlet series

$$\beta(s) := \prod_{p>2 \text{ prime}} (1 - p^{-s}) = \sum_{k=1}^{\infty} \mu_2(k) k^{-s}.$$

If μ is the Möbius function and $\zeta(s)^{-1} = \sum_{k=1}^{\infty} \mu(k)k^{-s} = \prod_{p \text{ prime}}(1 - p^{-s})$ is the reciprocal of the Riemann zeta function, then $\beta(s) = ((1-2^{-s})\zeta(s))^{-1}$, and μ_2 is described explicitly by $\mu_2(k) = 0$ if k is even and $\mu_2(k) = \mu(k)$ if k is odd.

Now, by the inclusion-exclusion principle, the left-hand side of (13) is equal to

$$\frac{1}{x^2} \sum_{k=1}^{\infty} \mu_2(k) \big[g_0(Ax;k) g_1(Bx;k) + g_1(Ax;k) g_0(Bx;k) \big].$$

Since clearly $|g_i(u;k) - (u/2k)| \le 1$ (for k odd), this is easily seen to be $cAB + O((\log x)/x)$ as $x \to \infty$, where

$$c = \frac{1}{2} \sum_{k=1}^{\infty} \mu_2(k) k^{-2} = \frac{1}{2}\beta(2) = \frac{1}{2} \cdot \frac{4}{3} \zeta(2)^{-1} = \frac{4}{\pi^2},$$

proving our claim.

It follows, by taking unions and differences, that (13) is true for D any finite union of rectangles with sides parallel to the coordinate axes, and therefore by approximation for any bounded and open D.

For $0 < t \le \pi/2$, denote

$$\begin{aligned} \mathrm{arc}(t) &= \Big\{ (\cos u, \sin u) : 0 < u < t \Big\}, \\ \mathrm{sector}(t) &= \Big\{ (x,y) : x, y > 0, \quad x^2 + y^2 \le 1, \quad \arctan(y/x) < t \Big\}. \end{aligned}$$

By the parametric solution (6) we have as $N \to \infty$

$$\begin{aligned} & \mathbb{P}_N^{oe}\Big((a,b,c) \in \mathrm{PPT}_N : (a/c, b/c) \in \mathrm{arc}(t) \Big) \\ = \ & \frac{\#\Big\{ (m,n) \in \mathbb{Z}^2 : \ m > n > 0, \ \gcd(m,n) = 1, \ m+n \equiv 1(2), \ m^2+n^2 \le N, \ \arctan\left(\frac{2mn}{m^2-n^2}\right) < t \Big\}}{\#\Big\{ (m,n) \in \mathbb{Z}^2 : \ m > n > 0, \ \gcd(m,n) = 1, \ m+n \equiv 1(2), \ m^2+n^2 \le N \Big\}} \\ = \ & \frac{\#\Big\{ (m,n) \in \mathbb{Z}^2 : \ \gcd(m,n) = 1, \ m+n \equiv 1(2), \ \left(\frac{m}{\sqrt{N}}, \frac{n}{\sqrt{N}}\right) \in \mathrm{sector}(t/2) \Big\}}{\#\Big\{ (m,n) \in \mathbb{Z}^2 : \ \gcd(m,n) = 1, \ m+n \equiv 1(2), \ \left(\frac{m}{\sqrt{N}}, \frac{n}{\sqrt{N}}\right) \in \mathrm{sector}(\pi/4) \Big\}} \\ = \ & (1+o(1)) \frac{4\pi^{-2} \cdot \mathrm{area}(\mathrm{sector}(t/2)) N}{4\pi^{-2} \cdot \mathrm{area}(\mathrm{sector}(\pi/4)) N} = (1+o(1)) \frac{2t}{\pi}. \end{aligned}$$

This is exactly the claim of the Lemma. ■

Define the *Perron-Frobenius operator* of T as the operator $H : L_1(\mathcal{Q}, \lambda) \to L_1(\mathcal{Q}, \lambda)$,

$$\begin{aligned}(Hf)(x,y) &= \frac{1}{3+2x-2y} \cdot f\left(\frac{2+x-2y}{3+2x-2y}, \frac{2+2x-y}{3+2x-2y}\right) \\ &+ \frac{1}{3+2x+2y} \cdot f\left(\frac{2+x+2y}{3+2x+2y}, \frac{2+2x+y}{3+2x+2y}\right) \\ &+ \frac{1}{3-2x+2y} \cdot f\left(\frac{2-x+2y}{3-2x+2y}, \frac{2-2x+y}{3-2x+2y}\right).\end{aligned}$$

H is also known as the *transfer operator* corresponding to T. It has the property that if the random vector (X, Y) on $\mathcal{Q}$ has distribution $f(x, y)d\lambda$, then $T(X, Y)$ has distribution $(Hf)(x, y)d\lambda$. We skip the simple computation that verifies this claim.

Theorem 8. *Under $\mathbb{P}_N$, the distribution of $d(T^{n-1}(a/c, b/c))$, the n-th digit in the expansion of a random PPT $(a, b, c) \in PPT_N$, converges to the distribution of $d(x, y)$ under the measure*

$$d\lambda_n(x,y) = (H^{n-1}(\mathbf{1}))(x,y)d\lambda(x,y),$$

where $\mathbf{1}$ is the constant function 1.

Proof. This is immediate from Lemma 7 and the definition of H. ■

Theorem 8 answers the question of the limiting distribution as $N \to \infty$ of the digits d_n in the expansion of a random PPT; however, it does not give good asymptotic information on the behavior of these distributions as $n \to \infty$. In fact, this is not a very interesting question: since the invariant measure is infinite, the density $H^n(\mathbf{1})$ will become for large n more and more concentrated around the singular points $(1, 0)$ and $(0, 1)$ (it is possible to make this statement more precise, but we do not pursue this slightly technical issue here).

Here's one way to amend the situation in a way that enables formulating interesting quantitative statements concerning the distribution of the digits, which we mention briefly without going into detail: replace the expansion $(d_k)_{k=1}^n$ by a new expansion $(e_j)_{j=1}^{\ell}$, by dividing the (d_k) into blocks consisting of 1's and 3's and terminating with a 2; so for instance, the expansion $(1, 1, 2, 2, 3, 1, 3, 2, 3, 3, 3, 1, 2, 1)$ will be replaced by $(112, 2, 3132, 33312, 1)$. The new expansion corresponds to the induced transformation T_B, where B is the middle arc in the generating partition. Cylinder sets of T_B can be easily computed. The invariant measure is the restriction $\mu_{|B}$, a *finite* measure, so normalize it to be a probability measure. T_B is easily shown to be exact as in section 3.3. A theorem analogous to Theorem 8 above can be proved, to the effect that the n-th digit in the "new" expansion of a random PPT converges in distribution to the distribution of $d_{\text{new}}(x, y)$ (the first "new" digit) under the measure whose density with respect to λ is the $(n-1)$-th iterate of the Perron-Frobenius operator of T_B applied to the constant function $\mathbf{1}$. Since T_B is mixing, these densities will actually converge to the invariant density. So after each occurrence of a "2" in the original expansion, there are well-defined statistics for the sequence of digits that follows up to the next "2", which can be computed easily by integrating the invariant density (5) over the appropriate cylinder set.

We conclude this section with some open problems: study the expectation, the variance and the limiting distribution of the *length* of the expansion of a random element of PPT_N, as $N \to \infty$. Generalize to arbitrary cost-functions, as in [8].

5 Concluding remarks

5.1 Some special expansions

From a number-theoretic standpoint, it is interesting to study points on $\mathcal{Q}$ with special expansions. As the examples in section 1 show, simple periodic expansions seem to correspond to simple quadratic points on $\mathcal{Q}$. It is easy to see that any eventually-periodic expansion corresponds to the image under D of a quadratic irrational. Do all quadratic irrationals have eventually periodic expansions?

We also found empirically the expansions

$$\begin{aligned} e^i &= (\cos 1, \sin 1) &= [3, 1^2, 3, 1^4, 3, 1^6, 3, 1^8, 3, \ldots], \\ e^{i/2} &= (\cos 1/2, \sin 1/2) &= [1, 3, 1^5, 3, 1^9, 3, 1^{13}, 3, \ldots]. \end{aligned}$$

where 1^k means a succession of k 1's. The first equation can be proved by observing that $D^{-1}(\cos 1, \sin 1) = (1-\cos(1))/\sin(1) = \tan(1/2)$, and that the approximations

$$\left(F_3 \circ F_1^2 \circ F_3 \circ F_1^4 \circ \ldots \circ F_3 \circ F_1^{2k}\right)(1)$$

$(F_1, F_2, F_3$ as in (11)) have the continued fraction expansions

$$\frac{1}{1}\,+\,\frac{1}{1}\,+\,\frac{1}{4}\,+\,\frac{1}{1}\,+\,\frac{1}{8}\,+\,\frac{1}{1}\,+\,\frac{1}{12}\,+\,\cdots\,+\,\frac{1}{4k}.$$

Thus our expansion reduces to the known ([26, sequence A019425]) infinite continued fraction expansion

$$\tan(1/2) = \frac{1}{1}\,+\,\frac{1}{1}\,+\,\frac{1}{4}\,+\,\frac{1}{1}\,+\,\frac{1}{8}\,+\,\frac{1}{1}\,+\,\frac{1}{12}\,+\,\cdots$$

The second equation is proved similarly. It is interesting to wonder whether other "nice" expansions exist for simple points on $\mathcal{Q}$.

5.2 Other number-theoretic expansions

While historically, the Pythagorean triples and their associated quadratic form $a^2+b^2-c^2$ are one of the oldest and most extensively studied Diophantine equations, one might wonder if matrix expansions such as the one in Theorem 1 and its associated dynamical system might exist for other Diophantine equations, and specifically for homogeneous ternary quadratic forms of signature $(2,1)$. Indeed, the case of the so-called 120-degree triples, namely triples of positive integers a, b, c which form the side lengths of a 120-degree triangle (and therefore satisfy the equation $a^2+ab+b^2=c^2$), was treated in [28] (see also [22]) where it was shown that the set of solutions can be generated (with uniqueness as in Theorem 1) from the two fundamental solution vectors $(3,5,7)^t$ and $(8,7,13)^t$ using the five generating matrices

$$N_1 = \begin{pmatrix} -1 & -4 & 4 \\ 1 & -3 & 4 \\ 0 & -6 & 7 \end{pmatrix}, \quad N_2 = \begin{pmatrix} -1 & 3 & 4 \\ 1 & 4 & 4 \\ 0 & 6 & 7 \end{pmatrix}, \quad N_3 = \begin{pmatrix} 4 & 3 & 4 \\ 3 & 4 & 4 \\ 6 & 6 & 7 \end{pmatrix},$$

$$N_4 = \begin{pmatrix} 4 & 1 & 4 \\ 3 & -1 & 4 \\ 6 & 0 & 7 \end{pmatrix}, \quad N_5 = \begin{pmatrix} -3 & 1 & 4 \\ -4 & -1 & 4 \\ -6 & 0 & 7 \end{pmatrix}.$$

It would be interesting to give an analysis of the dynamical system arising from this expansion. The underlying space would be the positive quadrant of the rotated ellipse $x^2 + xy + y^2 = 1$, and we expect much of the analysis in this paper to carry over to this case without difficulty. More ambitiously, one might try to classify the different possible expansions that can arise from ternary quadratic forms.

Acknowledgements

Thanks to Jon Aaronson and to Boaz Klartag for helpful discussions.

References

[1] J. Aaronson, An Introduction to Infinite Ergodic Theory. Mathematical Surveys and Monographs 50, Amer. Math. Soc, Providence, RI, 1997.

[2] J. Aaronson, M. Denker, The Poincaré series of $\mathbb{C} \setminus \mathbb{Z}$. *Ergodic Theory Dynam. Systems* 19 (1999), 1–20.

[3] R. L. Adler, Geodesic flows, interval maps, and symbolic dynamics. In: Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces, 93–123. Oxford University Press, Oxford, 1991.

[4] R. L. Adler, L. Flatto, Cross section maps for geodesic flows I (the modular surface). In: Ergodic theory and dynamical systems, vol. 2 (College Park, Maryland 1979-1980), 103–161. Progr. Math., 21, Birkhäuser, Boston, Mass., 1982.

[5] R. L. Adler, L. Flatto, The backward continued fraction map and geodesic flow. *Ergodic Theory Dynam. Systems* 4 (1984), 487–492.

[6] R. L. Adler, L. Flatto, Cross section map for the geodesic flow on the modular surface. In: Conference in modern analysis and probability (New Haven, Conn., 1982), 9–24. Contemp. Math., 26, Amer. Math. Soc., Providence, RI, 1984.

[7] R. Alperin, The Modular tree of Pythagoras. *Amer. Math. Monthly* 112 (2005), 807–816.

[8] V. Baladi, B. Valleé, Euclidean algorithms are Gaussian. J. Number Theory 110 (2005), 331–386.

[9] F. J. M. Barning, On Pythagorean and quasi-Pythagorean triangles and a generation process with the help of unimodular matrices. (Dutch) *Math. Centrum Amsterdam Afd. Zuivere Wisk.*, ZW-011 (1963).

[10] P. Billingsley, Ergodic Theory and Information. John Wiley & Sons, New York-London-Sydney 1965.

[11] J. Gollnick, H. Scheid, J. Zöllner, Rekursive Erzeugung der primitiven pythagoreischen Tripel. (German) *Math. Semesterber.* 39 (1992), 85–88.

[12] A. Hall, Genealogy of Pythagorean triads. *Math. Gazette* 54:390 (1970), 377–379.

[13] G. H. Hardy, E. M. Wright, An Introduction to the Theory of Numbers, 5th ed. Oxford University Press, Oxford, 1985.

[14] F. Herzog, B. M. Stewart, Patterns of visible and nonvisible lattice points. *Amer. Math. Monthly* 78 (1971), 487–496.

[15] J. Jaeger, Pythagorean number sets. (Danish.) *Nordisk Mat. Tidskr.* 24 (1976), 56–60, 75.

[16] A. R. Kanga, The family tree of Pythagorean triples. *Bull. Inst. Math. Appl.* 26 (1990), 15–17.

[17] D. E. Knuth, The Art of Computer Programming, vol. 2: Seminumerical Algorithms, 3rd. ed. Addison-Wesley, 1998.

[18] E. Kristensen, Pythagorean number sets and orthonormal matrices. (Danish) *Nordisk Mat. Tidskr.* 24 (1976), 111–122, 135.

[19] A. Lönnemo, The trinary tree underlying primitive pythagorean triples. In: Cut the Knot, Interactive Mathematics Miscellany and Puzzles. Alex Bogomolny (Ed.), `http://www.cut-the-knot.org/pythagoras/PT_matrix.shtml`.

[20] D. McCullough, Height and excess of Pythagorean triples. *Math. Mag.* 78 (2005), 26–44.

[21] P. Préau, Un graphe ternaire associé à l'équation $X^2 + Y^2 = Z^2$. *C. R. Acad. Sci. Paris Ser. I Math.* 319 (1994), 665–668.

[22] A. Redmon, Pythagorean triples, Master's Thesis, Eastern Washington University, 1998.

[23] F. Schweiger, Ergodic Theory of Fibred Systems and Metric Number Theory. Clarendon Press, Oxford, 1995.

[24] C. Series, The modular surface and continued fractions. *J. London Math. Soc.* (2) 31 (1985), 69–80.

[25] C. Series, Geometrical methods of symbolic coding. In: Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces, 125–151. Oxford University Press, Oxford, 1991.

[26] N. J. A. Sloane, editor (2003), The On-Line Encyclopedia of Integer Sequences. `http://www.research.att.com/ njas/sequences/`.

[27] B. Vallée, Dynamical analysis of a class of Euclidean algorithms. *Theor. Comp. Sci.* 297 (2003), 447–486.

[28] A. Wayne, A genealogy of 120^o and 60^o triangles. *Math. Mag.* 55 (1982), 157–162.